WHAT IS CLAIMED IS:

1. An automatic speech segmentation and verification method comprising:

a retrieving step, for retrieving a recorded speech corpus, the recorded speech corpus corresponding to a known text script, the known text script defining phonetic information with N phonetic units;

a segmenting step, for segmenting the recorded speech corpus into N test speech unit segments referring to the phonetic information of the N phonetic units in the known text script;

a segment-confidence-measure verifying step, for verifying segment confidence measures of N cutting points of the test speech unit segments to determine if the N cutting points of the test speech unit segments are correct;

a phonetic-confidence-measure verifying step, for verifying phonetic confidence measures of the test speech unit segments to determine if the test speech unit segments correspond to the known text script; and

a determining step, for determining acceptance of the phonetic unit by comparing a combination of segment reliability and the phonetic confidence measures of the test speech unit segments to a predetermined threshold value; wherein if the combined confidence measure is greater than the predetermined threshold value, the phonetic is accepted.

2. The method as claimed in claim 1, wherein the segmenting step further comprises:

using a hidden Markov model (HMM) to cut the recorded speech corpus into N test speech unit segments referring to the phonetic information of the N phonetic units in the known text script, wherein each test speech unit segment is defined as correspondingly having an initial cutting point;

performing a fine adjustment on the initial cutting point of the test speech unit segment according to at least one feature factor corresponding to each test speech unit segment and calculating at least one cutting point fine adjustment value corresponding to each test speech unit segment; and

integrating the initial cutting point and the cutting point fine adjustment value of the test speech unit segment to obtain a cutting point of the test speech unit segment.

3. The method as claimed in claim 2, wherein the feature factor of the test speech unit segment is a neighboring cutting point of the initial cutting point.

4. The method as claimed in claim 2, wherein the feature factor of the test speech unit segment is a zero crossing rate (ZCR) of the test speech unit segment.

5. The method as claimed in claim 2, wherein the feature factor of the test speech unit segment is an energy value of the test speech unit segment.

6. The method as claimed in claim 5, wherein the energy value is an energy value of a band pass signal and a high pass signal retrieved from a speaker-dependent band.

7. The method as claimed in claim 2, wherein each cutting point fine adjustment value has a weighted value, and the cutting point of the test speech unit segment is a weighted average of the initial cutting point and the cutting point fine adjustment value.

8. The method as claimed in claim 1, wherein in the segment-confidence-measure step, each segment confidence measure of the test speech unit segment is:

$$\mathrm{CMS} = \max\left(1 - \mathrm{h}(\mathrm{D}) - \sum_{s,f} \mathrm{g}(\mathrm{c}(s), \mathrm{f}(s)), 0\right),$$

where $h(D) = K\left(\sum_i w_i \left| d_i - \bar{d} \right|\right)$, D is a vector of multiple expert decisions of the cutting point, d_i is the cutting point, $\bar{d} = p(D)$ is a final decision of the cutting point, K(x) is a monotonically increasing function that maps a non-negative variable x into a value between 0 and 1, g(c(s), f(s)) is a cost function value between a cost function ranging from 0 to 1, s is a segment, c(s) is a type category of the segment s and, f(s) are acoustic features of the segment.

9. The method as claimed in claim 1, wherein in the phonetic-confidence-measure step, each phonetic confidence measure of the test speech unit segments is:

$$CMV = \min\{LLR_I, LLR_F, 0\},$$

where $\begin{cases} LLR_I = \log P(X_I \mid H_0) - \log P(X_I \mid H_1) \\ LLR_F = \log P(X_F \mid H_0) - \log P(X_F \mid H_1) \end{cases}$, X_I is an initial segment of the test speech unit segment, X_F is a final segment of the test speech unit segment, H_0 is a null hypothesis of the test speech unit segment recorded correctly, H_1 is an alternative hypothesis of the test speech unit segment recorded incorrectly, and LLR is a log likelihood ratio.

10. An automatic speech segmentation and verification system comprising:

a database for storing a known text script and a recorded speech corpus corresponding to the known text script, and the known text script has phonetic information with N speech unit segment wherein N is a positive integer;

a speech unit segmentor for segmenting the recorded speech corpus into N test speech unit segments referring to the phonetic information of the known text script;

a segmental verifier for verifying the correctness of the cutting points of test speech unit segments by obtaining a segmental confidence measure;

a phonetic verifier for obtaining a confidence measure of syllable verification by using verification models for verifying whether the recorded speech corpus is correctly recorded; and

a speech unit inspector for integrating the confidence measure of syllable segmentation and the confidence measure of syllable verification to determine whether the test speech unit segment is accepted.

11. The system as claimed in claim 10, wherein the segmental verifier performs the following steps:

using a hidden Markov model (HMM) to cut the recorded speech corpus into N test speech unit segments referring to the phonetic information of the N phonetic units in the known text script, wherein each test speech unit segment is defined as correspondingly having an initial cutting point;

performing a fine adjustment on the initial cutting point of the test speech unit segment according to at least one feature factor corresponding to each test speech unit segment and calculating at least one cutting point fine adjustment value corresponding to each test speech unit segment; and

integrating the initial cutting point and the cutting point fine adjustment value of the test speech unit segment to obtain a cutting point of the test speech unit segment.

12. The system as claimed in claim 11, wherein the feature factor of the test speech unit segment is a neighboring cutting point of the initial cutting point.

13. The system as claimed in claim 11, wherein the feature factor of the test speech unit segment is a zero crossing rate (ZCR) of the test speech unit segment.

14. The system as claimed in claim 11, wherein the feature factor of the test speech unit segment is an energy value of the test speech unit segment.

15. The system as claimed in claim 14, wherein the energy value is an energy value of a band pass signal and a high pass signal retrieved from a speaker-dependent band.

16. The system as claimed in claim 11, wherein each cutting point fine adjustment value has a weighted value, and the cutting point of the test speech unit segment is a weighted average of the initial cutting point and the cutting point fine adjustment value.

17. The system as claimed in claim 10, wherein each segment confidence measure of the test speech unit segment is determined by:

$$CMS = \max\left(1 - h(D) - \sum_{s,f} g(c(s), f(s)), 0\right),$$

where $h(D) = K\left(\sum_i w_i \left| d_i - \bar{d} \right|\right)$, D is the vector of multiple expert decisions of the cutting point, d_i is the cutting point, $\bar{d} = p(D)$ is a final decision of the cutting point, K(x) is a monotonically increasing function that maps a non-negative variable x into a value between 0 and 1, g(c(s), f(s)) is a cost function value between a cost function ranging from 0 to 1, s is a segment, c(s) is the type category of the segment s and, f(s) is the acoustic feature of the segment.

18. The method as claimed in claim 10, wherein each phonetic confidence measure of the test speech unit segments is determined by:

$$CMV = \min\{LLR_I, LLR_F, 0\},$$

where $\begin{cases} LLR_I = \log P(X_I \mid H_0) - \log P(X_I \mid H_1) \\ LLR_F = \log P(X_F \mid H_0) - \log P(X_F \mid H_1) \end{cases}$, X_I is initial segment of the test speech unit segment, X_F is final segment of the test speech unit segment, H_0 is a null hypothesis of the test speech unit segment recorded correctly, H_1

is an alternative hypothesis of the test speech unit segment recorded incorrectly, and LLR is a log likelihood ratio.